Optimal Sampling Strategies for Oceanic Applications

Dr Peter R. Oke
CSIRO Marine and Atmospheric Research and Wealth from Oceans Flagship Program
GPO Box 1538 Hobart TAS 7001 Australia
phone: (+61 3) 6232-5387   fax: (+61 3) 6232-5123   email: peter.oke@csiro.au

LONG-TERM GOALS

The long-term goals of this project are to improve our ability to monitor and predict the ocean circulation through the assessment and design of ocean observing systems; and through the development of practical methods for ocean data assimilation.

OBJECTIVES

The immediate scientific objective of this research is to develop a suite of ensemble-based data assimilation tools for the objective design and assessment of ocean observing systems for long-term monitoring programs.

APPROACH

The methods developed under this project exploit ensemble-data assimilation theory. They build on the work of Bishop et al. (2001). The methods seek to identify the set of observations that minimise the analysis error variance for a predefined variable or quantity (e.g., surface currents in a given region).

Under this project, the development and assessment of new methods in data assimilation and observing system design is initially conducted through their application to small, idealised models (e.g., Lorenz and Emmanuel 1998; Oke et al. 2006). This approach facilitates a thorough examination of the theoretical properties, limitations and strengths of different approaches.

Subsequent to their development and testing on idealised models, the techniques investigated under this project are applied to realistic applications to support the design and maintenance of ongoing and planned observation programs. This includes assessments of the Global Ocean Observing System (GOOS) for constraining an eddy-resolving ocean forecast system (e.g., Oke and Schiller 2007); and the design and assessment of various components of the Australian Integrated Marine Observing System (IMOS; see related projects). For these activities, we exploit the Bluelink ocean forecast and reanalysis system, developed by CSIRO and the Bureau of Meteorology (BoM). The Bluelink system was developed and tested through a 15-year reanalysis, called the Bluelink ReANalysis (BRAN; Oke et al. 2008a; Schiller et al. 2008. The Bluelink forecast system became operational at the BoM in August 2007 (www.bom.gov.au/oceanography/forecasts).

1. REPORT DATE **30 SEP 2008**	2. REPORT TYPE **Annual**	3. DATES COVERED **00-00-2008 to 00-00-2008**
4. TITLE AND SUBTITLE **Optimal Sampling Strategies For Oceanic Applications**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **CSIRO Marine and Atmospheric Research and Wealth from Oceans Flagship Program,GPO Box 1538 Hobart TAS 7001 Australia, ,**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
 Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
 code 1 only

14. ABSTRACT
The long-term goals of this project are to improve our ability to monitor and predict the ocean circulation through the assessment and design of ocean observing systems; and through the development of practical methods for ocean data assimilation.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**9**	

Dr. Peter Oke is the P.I. on this project and leads the data assimilation activities at the Centre for Australian Climate and Weather Research (CAWCR), a partnership between CSIRO and the BoM. Other researchers contributing to this research include Dr Pavel Sakov, now based at NERSC in Norway, and Dr. Hans Ngodock and Dr. Gregg Jacobs from NRL.

WORK COMPLETED

We have developed a method for objective array design and applied it to a suite of applications, including the design of a surface mooring array in the tropical Indian Ocean (Sakov and Oke 2008a). Subsequently, we have generalized and tested the method for multivariate applications and to represent different observation types that include satellite observations of sea-level anomaly (SLA) and sea-surface temperature (SST), in situ observations of temperature (T), salinity (S), sea-level and velocity (u and v), and land-based radar observations of u and v. This capability has been applied to the design and assessment of a component of the Australian IMOS. This includes an evaluation of different options for the deployment of HF radars and gliders off the south-east coast of Australia (Oke et al. 2008b).

To date, all of the techniques employed under this project have been based on ensemble data assimilation theory. In the course of this research we have explored the importance of different assimilation algorithms for ensemble square root filters (ESRFs). Based on these explorations, we have shown that only mean-preserving ensemble transformations should be used. These findings have been documented by Sakov and Oke (2008b). We have also proposed a new flavour of the EnKF that we call the deterministic EnKF (DEnKF). The DEnKF provides a computationally efficient alternative formulation to other ESRFs, but unlike most ESRFs, the DEnKF readily permits covariance localisation, making it an attractive option for realistic, large dimension applications. The details of the DEnKF are documented by Sakov and Oke (2008c).

A series of observing system experiments (OSEs) has been undertaken to investigate the relative importance of different observation types (Argo, altimeter and SST data) for constraining an eddy resolving ocean model (Oke and Schiller 2007). This work has recently been extended, using an analysis of the so-called influence matrix (Cardinali et al. 2004) to evaluate the influence of individual observations in a given analysis. This approach is inexpensive; and can readily be applied in an operational environment to monitor and assess the GOOS in near-real-time.

RESULTS

The objective of the array design method described by Sakov and Oke (2008a) is based on ensemble data assimilation theory. This method simply takes a background ensemble A^b, from an EnKF, ESRF or Ensemble Optimal Interpolation (EnOI) system and estimates the analysis ensemble A^a that results from assimilating a set of observations, defined by H and R. Here H is the observation operator and R is the observation error covariance matrix. In practice, R is estimated based on the expected errors associated with a hypothetical set of observations and H is constructed to represent the details of the observing system, including the observation types (i.e., what is to be observed) and the observation locations. The update of A^b to A^a is achieved using a matrix transformation that is borrowed from ESRF theory (Bishop et al. 2001). We examine the diagonals of the ensemble estimates of the background error covariance matrix P^b, and the analysis error covariance matrix P^a, given by:

$$P^b = A^b A^{bT} / (m\text{-}1), \text{ and} \qquad\qquad (1)$$

$$P^a = A^a A^{aT} / (m\text{-}1) \qquad\qquad (2)$$

where m is the ensemble size, to quantify the percentage improvement $\%I$, achieved by assimilating a given set of observations:

$$\%I = (\varepsilon_{STD} - \varepsilon_{STD+NEW}) / \varepsilon_{STD} \; x \; 100, \qquad\qquad (3)$$

where ε_{STD} is the estimated standard deviation of the analysis errors when a standard array of observations are assimilated (e.g., altimetry, SST and Argo); and $\varepsilon_{STD+NEW}$ is the estimated standard deviation of the analysis errors when a new array of observations (e.g., HF radar) are added to the standard array of observations.

We have applied this method to a number of different scenarios including the design of a tropical Indian Ocean mooring array (Sakov and Oke 2008a) and the design of a shelf observation array off south-eastern Australia (Oke et al. 2008b).

For the latest study (Oke et al. 2008b), we have adopted the error statistics used by the Bluelink reanalysis and forecast system (Oke et al. 2008a). The Bluelink data assimilation system is based on EnOI. We simply use the Bluelink ensemble to quantify the background field errors and covariances in (1). The details of the Bluelink assimilation system are mimicked as closely as possible enabling an assessment of the likely benefits of future observations programs to the Bluelink forecast and reanalysis system. As a first step to this study, we assess the validity of the assumed error variances used under Bluelink (Figure 1). This involves a comparison between the background field (BGF) and analysis
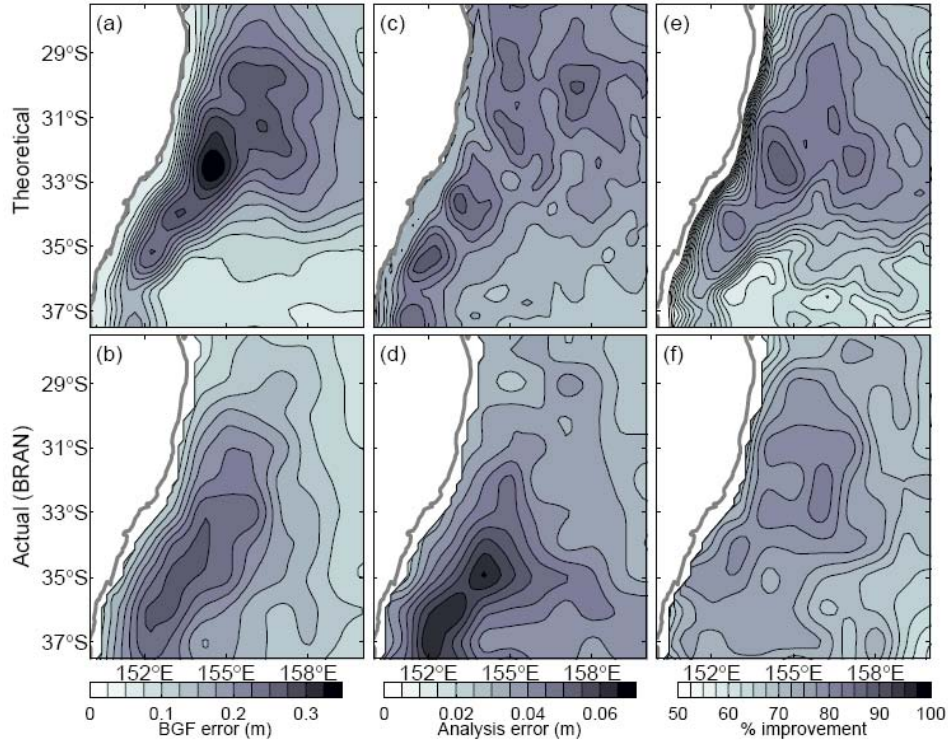
Figure 1: Maps showing the theoretical (top) and actual (bottom) BGF errors (a-b), analysis errors (c-d) and percentage improvement (e-f), from (7), for SLA. Contour intervals for the BGF error, analysis error and percentage improvement are 0.03 m, 0.005 m and 2.5% respectively. Note that the range is different for the BGF and analysis errors. Adapted from Oke et al. 2008b, submitted. [Figure showing comparisons between the assumed errors used by Bluelink and the actual errors based on a long reanalysis]

errors assumed by the Bluelink system (labeled as "Theoretical" in Figure 1) and the actual errors from a long data assimilating integration (labeled "Actual (BRAN)" in Figure 1). The actual errors presented in Figure 1 are estimated by comparing model fields and analyses to along-track observations from altimetry. If the errors assumed by the Bluelink Ocean Data Assimilation System (BODAS; Oke et al. 2008a) are correct, the fields in the top and bottom rows of each column of Figure 1 would be equivalent. They are clearly different in detail, but show many similar features, like the band of high error aligned with the coast. Similarly, the order of magnitude of these fields are similar for both the theoretical and actual estimates (BGF errors are 0.1-0.3 m; Analysis errors are 0.03-0.06 m). These comparisons indicate that the error estimates used in Bluelink, and in the observing system study described here, are imprecise, but provide a reasonable first guess.

Subsequent to the evaluation of the error statistics used under Bluelink, we have undertaken a study to assess the likely benefits of the component of the Australian IMOS that is planned for the continental shelf waters off New South Wales, Australia, hereafter, the NSW-IMOS. Under NSW-IMOS, the deployment of one or two short-range HF radar arrays is planned, along with the deployment of a Slocum glider. There are several locations where the HF radar systems could possibly be deployed. We estimate the expected benefits of the different options to the Bluelink system. We consider both short-range and long-range HF radars. We also estimate the impact of assimilating T and S observations from gliders deployed along sections of constant latitude within 200 km of the coast. Results showing

the %I, from equation (3), for the HF radar assessment, based on 50 independent realisations of the analysis errors, are shown in Figure 2. We find that, as expected, assimilation of HF radar observations should significantly improve the surface velocity fields in the vicinity of the observations. But we also find that assimilation of HF radar observations should have a significant benefit to other model variables, including sub-surface T and S, and sea-level. Similar results for gliders suggest that while the benefits of assimilating the glider observations are likely to be limited to the immediate vicinity of the observations, owing to the short decorrelation length-scales over the continental shelf, the analysis errors near the observations are likely to be significantly reduced by assimilating the glider data.

Common to the application of most statistical data assimilation systems, such as OI, EnOI, EnKF, ESRF and 3DVar is the employment of a gain matrix, $\mathbf{K}$. The gain matrix depends on the observation and background error covariances assumed by the system. A simple function of the gain matrix, referred to here as the influence matrix, is given by $\mathbf{K}^T\mathbf{H}^T$ (Cardinali et al. 2004). The influence matrix is a well-known concept in statistical fields, but has only recently been derived for modern data assimilation systems. The influence matrix can diagnose the relative contributions of observations and background fields to a given analysis; the relative influence of different observation types; and the effective degrees of freedom of signal (DFS) in different data types.

Using the practical method of Chapnik et al. (2006) for approximating the diagonal elements of the influence matrix, known as the self-sensitivities, we present some preliminary estimates of the Information Content (IC) and the DFS in Figure 3. The number of observations of each data type shown in Figure 3 refers to the number of super-observations in the Australian region (see Oke et al. 2008a for details). The DFS is the estimated sub-trace of $\mathbf{K}^T\mathbf{H}^T$ for each data type, and provides an estimate of the effective number of independent observations assimilated. The IC, presented as a percentage, is simply based on the ratio of the DFS and the number of observations. Based on these results, it is clear that both altimetry and SST observations are well used by the Bluelink system. However, information from the Argo data is clearly not extracted by the Bluelink system in an optimal way. This suggests that the error estimates used by the Bluelink system (Oke et al. 2008) need to be revisited and refined.
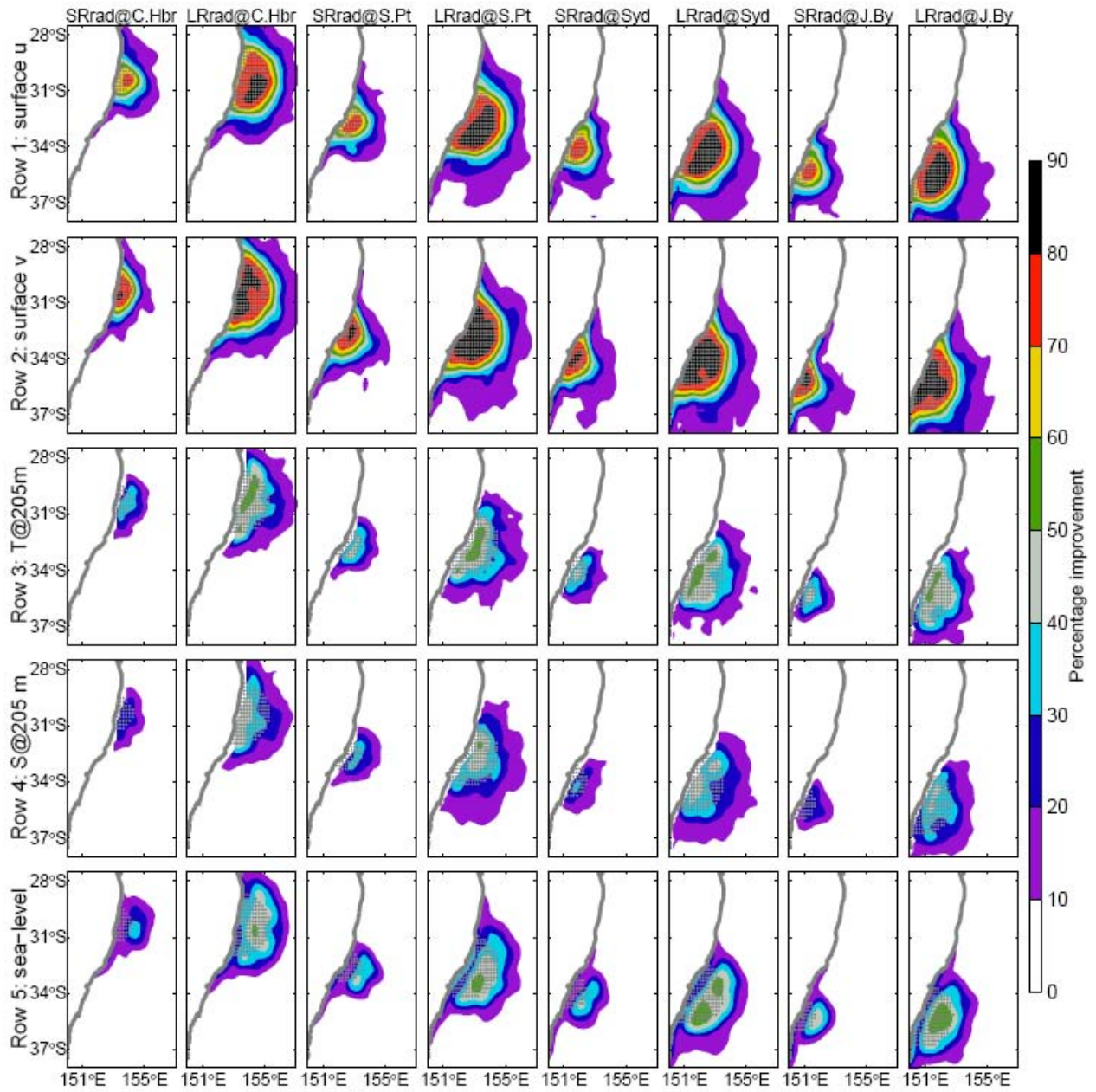
Figure 2: Estimated %I, from (3), for velocity (row 1-2), T and S at 205 m depth (row 3-4) and sea-level (row 5) for different options along the NSW coast. The locations of the assumed u and v observations are indicated in each column by the gray dots. Adapted from Oke et al. 2008b, submitted.
[Figure showing the expected % improvement in analyses when observations from different HF radar deployments are assimilated by the Bluelink]
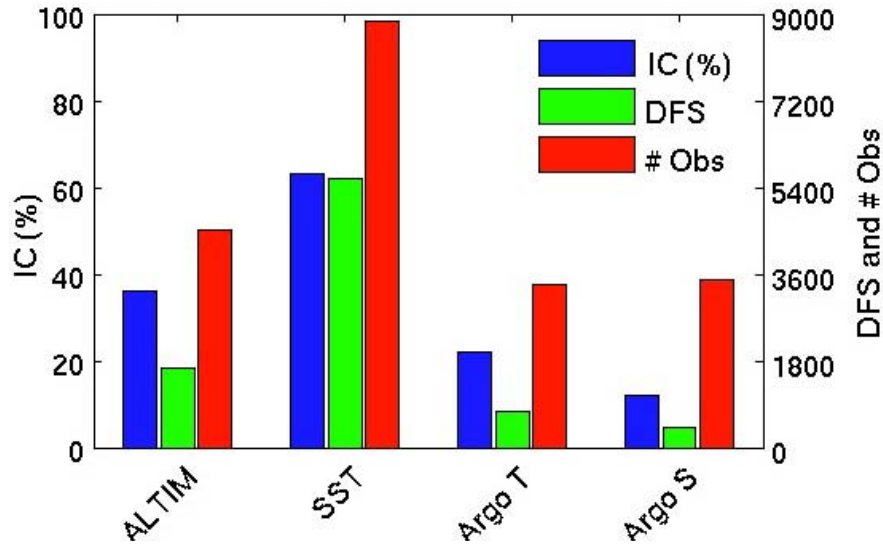
Figure 3: Preliminary estimates of the Information Content (IC; %), degrees of freedom of signal (DFS) and the number of assimilated super-observations (# Obs) for the Bluelink reanalysis system in the region 90-180°E, 60°S-equator, computed for 1 January 2006. The scale for the IC is to the left and the scale for the DFS and # Obs is to the right.
[Figure showing the information content of different observation types based on the error estimated used by the Bluelink system]

IMPACT/APPLICATIONS

The tools for objective array design that have been developed under this project are potentially very powerful. Given a time series, or model ensemble, of oceanic fields for some region, these tools facilitate the efficient design and assessment of an observation array for that region. Our experiments indicate that the most critical aspect of any application is formulating the cost function. That is, in determining exactly what it is that we wish to monitor. For example, an "optimal" array for monitoring intraseasonal mixed layer depth is likely to be very different from an "optimal" array for monitoring interannual variability (Sakov and Oke 2008a). The tools developed under this project have been applied to support the design of the NSW-IMOS, a new integrated marine observing system for south-eastern Australia (Oke et al. 2008a). The results described by Oke et al. (2008b), relating to the NSW-IMOS, are being used by the Bluelink science team to set priorities on the uptake on new observation types into the Bluelink forecast and reanalysis system.

The results from the OSEs presented by Oke and Schiller (2007) provide an important assessment of the performance of the Global Ocean Observing System (GOOS) for a data-assimilating, eddy-resolving ocean forecast system. Such information is essential for planning investment in the maintenance and development of the GOOS. The analyses of the influence matrix, relevant to the Bluelink system, provide guidance for refining the Bluelink assimilation system and improving its predictive skill.

Sakov and Oke (2008b) demonstrate, from both theory and a suite of applications, that only mean-preserving, or zero-centered, ensemble transformations should be used by ESRFs. This is a very clear

result, based on both theoretical and experimental evidence. This development has been adopted by the EnKF community (see http://enkf.nersc.no/Code/Analysis/meanpres.pdf).

RELATED PROJECTS

Bluelink is a partnership between CSIRO, the Bureau of Meteorology and the Royal Australian Navy. Many of the research activities undertaken in Bluelink have strong synergies for the project that is the subject of this annual report. The main objective of Bluelink is the development and application of an ocean forecast system for the mesoscale circulation around Australia. Applications of the Bluelink system are well documented (e.g., Oke et al. 2005; 2008a; Schiller et al. 2008).

The Australian Integrated Marine Observing System (IMOS) program (www.imos.org.au). IMOS involves the provision of observational platforms (e.g., gliders, high-frequency radars, moorings) to establish a long-term monitoring capability for the oceans around Australia.

REFERENCES

Bishop, C., B. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. part 1: theoretical aspects. *Monthly Weather Review*, **129**, 420–436.

Cardinala, C., S. Pezzulli and E. Andersson, 2004: Influence-matrix diagnostic of a data assimilation system. *Quarterly Journal of the Royal Meteorological Society,* **130**, 2767-2786.

Chapnik, B., G. Desroziers, F. Rabier and O. Talagrand, 2006: Diagnosis and tuning of observational error in a quasi-operational data assimilation setting. *Quarterly Journal of the Royal Meteorological Society,* **132**, 543-565.

Lorenz, E. N., and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *Journal of Atmospheric Sciences,* **55**, 399-414.

Oke, P. R., and A. Schiller, 2007: Impact of Argo, SST and altimeter data on an eddy-resolving ocean reanalysis. *Geophysical Research Letters,* **34**, L19601, doi:10.1029/2007GL031549.

Oke, P. R., A. Schiller, D. A. Griffin and G. B. Brassington, 2005: Ensemble data assimilation for an eddy-resolving ocean model, *Quarterly Journal of the Royal Meteorology Society,* **131**, 3301-3311.

Oke, P. R., P. Sakov, S. P. Corney, 2006: Impacts of localisation in the EnKF and EnOI: Experiments with a small model, Ocean Dynamics, http://dx.doi.org/10.1007/s10236-006-0088-8.

Oke, P. R., G. B. Brassington, D. A. Griffin and A. Schiller, 2008a: The Bluelink Ocean Data Assimilation System (BODAS). *Ocean Modelling,* **20**, 46-70.

Oke, P. R., P. Sakov and E. Schulz, 2008b: A comparison of shelf observation platforms for assimilation into an eddy-resolving ocean model. *Dynamics of Atmospheres and Oceans* submitted.

Sakov, P., and P. R. Oke 2008a: Objective array design: Application to the tropical Indian Ocean. *Journal of Atmospheric and Oceanic Technology,* **25**, 794-807.

Sakov, P., and P. R. Oke 2008b: Implications of the form of the ensemble transformations in the ensemble square root filters. *Monthly Weather Review,* **136**, 1042-1053.

Sakov, P. and P. R. Oke 2008c: A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus-A,* **60A**, 361-371.

PUBLICATIONS

Oke, P. R., P. Sakov and E. Schulz, 2008: A comparison of shelf observation platforms for assimilation into an eddy-resolving ocean model. *Dynamics of Atmospheres and Oceans* submitted. [refereed]

Oke, P. R., G. B. Brassington, D. A. Griffin and A. Schiller, 2008: The Bluelink Ocean Data Assimilation System (BODAS). *Ocean Modelling,* **20**, 46-70. [published, refereed]

Oke, P. R., and P. Sakov 2008: Representation error of oceanic observations for data assimilation. *Journal of Atmospheric and Oceanic Technology,* **25**, 1004-1017. [published, refereed]

Sakov, P., and P. R. Oke 2008: Objective array design: Application to the tropical Indian Ocean. *Journal of Atmospheric and Oceanic Technology,* **25**, 794-807. [published, refereed]

Sakov, P., and P. R. Oke 2008: Implications of the form of the ensemble transformations in the ensemble square root filters. *Monthly Weather Review,* **136**, 1042-1053. [published, refereed]

Sakov, P. and P. R. Oke 2008: A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus-A,* **60A**, 361-371. [published, refereed]

Schiller, A., P. R. Oke, G. B. Brassington, M. Entel, R. Fiedler, D. A. Griffin, and J. V. Mansbridge, 2008: Eddy-resolving ocean circulation in the Asian-Australian region inferred from an ocean reanalysis effort. *Progress in Oceanography,* **76**, 334-365. [published, refereed]

Turner, M. R. J., J. P. Walker, and P. R. Oke 2008: Ensemble member generation for sequential data assimilation. *Remote Sensing of Environment*, **112**, 1421-1433. [published, refereed]

Oke, P. R., and A. Schiller, 2007: Impact of Argo, SST and altimeter data on an eddy-resolving ocean reanalysis. *Geophysical Research Letters,* **34**, L19601, doi:10.1029/2007GL031549. [published, refereed]

Oke, P. R., and A. Schiller, 2007: A model-based assessment and design of a tropical Indian Ocean mooring array, *Journal of Climate,* **20**, 3269-3283. [published, refereed]

Preprints of submitted manuscripts are available from http://www.cmar.csiro.au/staff/oke/.

HONORS/AWARDS/PRIZES

Dr Peter Oke received a *Julius Career Award* from CSIRO. The *Julius Career Award* is designed to enhance the careers of early to mid-career scientists; and is intended to contribute towards their professional development.

Together with the Bluelink team, Dr Peter Oke was a finalist for the Eureka Prize for Outstanding Science in Support of Defense or National Security. Presented annually by the Australian Museum, *Eureka prizes* reward excellence in the fields of research and innovation, science leadership, school science and science journalism and communication.